

Basic Concepts

Liane Guillou and Alexander Fraser
{liane,fraser}@cis.uni-muenchen.de

CIS, Ludwig-Maximilians-Universität München

Morphology

2016

23.05.2016

Outline

1. Introduction
2. Morphological relationships
3. Morphological building blocks
4. Allomorphy

Slides adapted from Weller and Haselbach (IMS Stuttgart)

Originally based on: Haspelmath, M. & Sims, A. D. (2010): *Understanding Morphology* [2nd ed.], chapter 2 'Basic concepts', London: Hodder Education.

Outline

1. Introduction
2. Morphological relationships
3. Morphological building blocks
4. Allomorphy

Introduction

Segmentation

- Many words can be easily **segmented**,
i.e. broken up into individually meaningful parts
 - *read* *read-s* *read-er* *read-able*
wash *wash-es* *wash-er* *wash-able*
write *write-s* *writ-er* *writ-able*
 - *kind* *kind-ness* *un-kind*
happy *happi-ness* *un-happy*
friend-ly *friend-li-ness* *un-friend-ly*
- These meaningful parts are called **morphemes**
- Morphemes are the ultimate elements of morphological analysis; they are, so to speak, morphological atoms

Introduction

Morphemes

- **Morphemes** can be defined as the smallest meaningful constituents of a linguistic expression

- Example:

Camilla met an unfriendly chameleon.

- Possible segmentations:

- syntactic segmentation:

Camilla | met | an | unfriendly | chameleon.

- syntactic and morphological segmentation:

Camilla | met | an | un|friend|ly | chameleon.

- Impossible segmentation:

- *Camilla | met | an | un|friend|ly | *cha|meleon.*

Neither *cha* or *meleon* are meaningful in isolation, nor do they share any aspect of meaning in other contexts, e.g. **cha|risma*

Introduction

What is a word?

- Contiguous sequence of letters (i.e. sounds).
- How many words are in this sentence?

*Our plan was to meet there in a few days' time,
once our projects in Paris were concluded.*

(adapted from BNC)

- It depends on how you count.
 - 18 sequences of letters separated by blank spaces: **word tokens**
 - 16 different seq. of letters separated by blank spaces: **word types**
(*our* and *in* occur twice)
 - 15 different 'dictionary words': **lexemes**
(*{was,were}* → BE [verb lexeme])

Introduction

Contiguity of letters as a criterion for wordhood?

- In agglutinative languages, e.g. Turkish, 'words' can correspond to sentences rather than to lexemes

Ev-de-ydi-m.

home-LOC-PAST-1SG

'I was at home.'

Agglutination: complex words are formed by stringing together morphemes without changing their spelling or phonetics

Outline

1. Introduction
2. Morphological relationships
3. Morphological building blocks
4. Allomorphy

Morphological relationships

Lexeme

- A **lexeme** is a word in an abstract sense
- For example, the lexeme LIVE represents the core meaning shared by forms such as *live*, *lives*, *lived*, *living*
- In most languages, dictionaries are organised according to lexemes (cf. 'dictionary word')

Morphological relationships

Word-form

- A **word-form** is a word in a concrete sense. It is the combination of a lexeme and a set of grammatical meanings (or grammatical functions)
- For example the lexeme LIVE together with the grammatical meanings “third person, singular, present tense” yields the word-form *lives*
- Word-forms belonging to the same lexeme express different grammatical meanings, but the same core (semantic) concept
- **Test:** List some word-forms for the lexeme: READ

Morphological relationships

Paradigm

- The set of word-forms that belong to a lexeme is often called a **paradigm**
- Paradigm of the Modern Greek noun lexeme FILOS ('friend'):

	singular	plural
nominative	<i>fílos</i>	<i>fíli</i>
accusative	<i>fílo</i>	<i>fílus</i>
genitive	<i>fílu</i>	<i>fílon</i>

- Paradigm of the Latin noun INSULA ('island');
(note: 7 different sequences of sound, but 10 word-forms)

	singular	plural
nominative	<i>insula</i>	<i>insulae</i>
accusative	<i>insulam</i>	<i>insulās</i>
genitive	<i>insulae</i>	<i>insulārum</i>
dative	<i>insulae</i>	<i>insulīs</i>
ablative	<i>insulā</i>	<i>insulīs</i>

Morphological relationships

Word family

- Different lexemes may also be related to each other. A set of related lexemes is sometimes called a **word family**
- Two English word families:
 - READ, READABLE, UNREADABLE, READER, READABILITY, REREAD
 - LOGIC, LOGICIAN, LOGICAL, ILLOGICAL, ILLOGICALITY
- **Test:** suggest a set of lexemes that constitute a word family

Morphological relationships

Word family

- Each member of a word family is given its own dictionary entry
 - Complex lexemes generally denote new concepts that are different from the concepts of the corresponding simple lexemes (e.g. *read* denotes activity, *reader* denotes individual)
 - Complex lexemes are normally less predictable than word-forms (e.g. a specialist in logic is a *logician* rather than a **logicist*)
- In a word family there may be different parts of speech (V, N, ...), this cannot be the case within a paradigm

Morphological relationships

Inflection and derivation

- **Inflection** (= inflectional morphology):
The relationship between word-forms of a lexeme
- A lexeme inflects for (or: is inflected for) grammatical features, e.g. the Latin lexeme INSULA inflects for case and number
- **Derivation** (= derivational morphology):
The relationship between lexemes of a word family
- A lexeme can derive from (or: can be derived from) another lexeme, e.g. the lexeme READER is derived from the lexeme READ

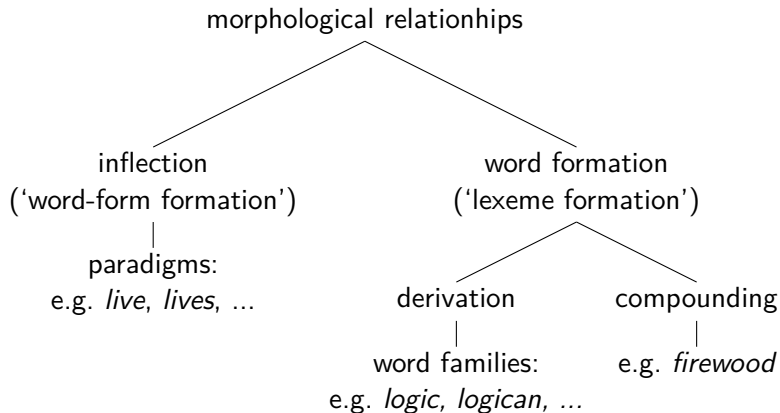
Morphological relationships

Compounding

- Some morphologically complex words belong to two or more word families simultaneously
- For example, the lexeme FIREWOOD belongs both in the word family of FIRE and in the word family of WOOD
- Such relationships are called **compounding** and the resulting lexemes are called **compound lexemes** or just **compounds**

Morphological relationships

Subdivisions of Morphology



Outline

1. Introduction
2. Morphological relationships
3. Morphological building blocks
4. Allomorphy

Morphological building blocks

Abstractness of meaning of morphemes

- In both inflection and derivation, morphemes have various kinds of meanings
- Some meanings are very **concrete** and can be described easily, e.g. the meanings of the morphemes *wash*, *logic*, *chameleon*, *un-*
- Other meanings are **abstract** and more difficult to describe, e.g. *-ity* in *readabil-ity* can be said to mean 'the *quality* of being readable'
- Some meanings are so abstract that they can hardly be called meanings, e.g. *-s* in English *read-s* is needed when the subject is a third person singular noun phrase. However it is difficult to say what the meaning of *-s* is. Instead we say that these kinds of morphemes have certain *grammatical functions*

Morphological building blocks

Affix and base

- Word-forms in an inflectional paradigm generally share (at least) one longer morpheme with a concrete meaning
- An **affix** attaches to a word or a main part of a word, i.e. to the **base**. The affix usually has an abstract meaning and cannot occur by itself.
- Russian case inflection (sg. forms, *ruk-* 'hand')

nominative	<i>ruk-a</i>
accusative	<i>ruk-u</i>
genitive	<i>ruk-i</i>
dative	<i>ruk-e</i>
locative	<i>ruk-e</i>
instrumental	<i>ruk-oj</i>

→ affix indicates case

Morphological building blocks

Affix and base

- Classical Nahuatl possessor inflection

1st sg.	<i>no-cal</i>	'my house'
2nd sg.	<i>mo-cal</i>	'your (SG) house'
3rd sg.	<i>i-cal</i>	'her/his house'
1st pl.	<i>to-cal</i>	'our house'
2nd pl.	<i>amo-cal</i>	'your (PL) house'
3rd pl.	<i>in-cal</i>	'their house'

→ affix indicates possessor

Morphological building blocks

Types of affixes

- Affixes can be characterised by their position within the word
- Types of affixes

Suffix	follows the base	Russian <i>-a</i> in <i>ruk-a</i> 'hand' English <i>-ful</i> in <i>event-ful</i>
Prefix	precedes the base	Classical Nahuatl <i>no-</i> in <i>no-cal</i> 'my house' English <i>un-</i> in <i>un-happy</i>
Infix	occurs inside the base	Arabic <i>-t-</i> in <i>(i)š-t-ağala</i> 'be occupied' (base: <i>šağala</i>) Tagalog <i>-um-</i> in <i>s-um-ulat</i> 'write' (base: <i>sulat</i>) Also: English swear words
Circumfix	occurs on both sides of the base	German <i>ge...-t</i> in <i>ge-mach-t</i> 'made' (base: <i>mach</i>)

Morphological building blocks

Test

- Suggest some affixes to attach to the base “like”
- Give positive and negative examples

Morphological building blocks

Root

- The **base** is a relative notion that is defined wrt. the notion **affix**
- A base that cannot be analyzed any further into constituent morphemes is called a **root**
 - E.g. Readable → read (root and base)
 - E.g. Readability → read (root); readable (base)

Morphological building blocks

Stem

- A base is also sometimes called a **stem**, especially in inflection
- A base may or may not be able to function as a word-form
 - Bases that can also function as word-forms are called **free stems**, e.g. *cat* is both the base of the inflected word-form *cats* and itself a word-form
 - Bases that cannot also function as word-forms are called **bound stems**, e.g. the Italian word-form *gatti* can be broken up into the plural suffix *-i* and the base *gatt-*, but both aren't word-forms by themselves. Italian nouns inflect must for number, even in the singular (e.g. *gatt-o* 'cat', *gatt-i* 'cats').

Morphological building blocks

Borderline cases

- Roots and affixes can generally be distinguished quite easily, but sometimes there are problems
- For example, the morphemes *bio-* and *-crat* could be regarded as affixes because they do not occur as independent lexemes, but their very concrete meanings suggest that they should be regarded as bound stems (i.e. roots) that only occur in compounds
 - *biogeography, bioethics, bioengineering, biorhythm, bioterrorism, biomedicine, biochip, etc.*
 - *artistocrat, autocrat, democrat, Eurocrat, plutocrat, technocrat, theocrat, etc.*

Outline

1. Introduction
2. Morphological relationships
3. Morphological building blocks
4. Allomorphy

Allomorphy

Allomorphs

- Morphemes may have different phonological shapes under different circumstances: **allomorph**
- For example, the plural morpheme/affix in English -s has various pronunciations:
 - [s] as in *cats* [kæts]
 - [z] as in *dogs* [dɒgz]
 - [əz] as in *faces* [feɪsəz]
- Roots can also have various shapes:
 - Rule: [i:] → [ɛ] (present-tense → past-tense form)
 - Instances: *sleep/slep-t*, *keep/kep-t*, *feel/fel-t*, *mean/mean-t*
- Allomorphs of one morpheme occur in different environments in **complementary distribution**. E.g. indefinite articles *a* and *an*:
 - an aardvark / * an bear
 - * a aardvark / a bear

Allomorphy

Examples of affixes

- Korean accusative suffix (marker of direct object): two allomorphs

<i>-ul</i>	<i>ton</i>	'money'	<i>ton-ul</i>	'money-ACC'
	<i>chayk</i>	'book'	<i>chayk-ul</i>	'book-ACC'
<i>-lul</i>	<i>tali</i>	'leg'	<i>tali-lul</i>	'leg-ACC'
	<i>sakwa</i>	'apple'	<i>sakwa-lul</i>	'apple-ACC'

(*-ul* after a consonant; *-lul* after a vowel)

Allomorphy

Examples of affixes

- Turkish first person possessive suffix: five allomorphs

<i>-im</i>	<i>ev</i>	'house'	<i>ev-im</i>	'my house'
	<i>dil</i>	'language'	<i>dil-im</i>	'my language'
<i>-üm</i>	<i>köy</i>	'village'	<i>köy-üm</i>	'my village'
	<i>gün</i>	'day'	<i>gün-üm</i>	'my day'
<i>-um</i>	<i>yol</i>	'way'	<i>yol-um</i>	'my way'
	<i>tuz</i>	'salt'	<i>tuz-um</i>	'my salt'
<i>-ım</i>	<i>ad</i>	'name'	<i>ad-ım</i>	'my name'
	<i>kız</i>	'girl'	<i>kız-ım</i>	'my daughter'
<i>-m</i>	<i>baba</i>	'father'	<i>baba-m</i>	'my father'

Allomorphy

Examples of roots

- German: a voiced obstruent (speech sound formed by obstructing air flow [consonants]) becomes voiceless (does not apply) in syllable-final position

<i>Tag</i>	[ta:k]	'day'	<i>Tage</i>	[ta:gə]	'days'
<i>Hund</i>	[hʊnt]	'dog'	<i>Hunde</i>	[hʊndə]	'dogs'
<i>Los</i>	[lo:s]	'lot'	<i>Lose</i>	[lo:ze]	'lots'

Allomorphy

Examples of roots

- Russian: when the stem is followed by a vowel-initial suffix, the vowel *o/e* is often dropped if it is the last vowel in the stem

<i>zamok</i>	'castle'	<i>zamk-i</i>	'castles'
<i>kamen'</i>	'stone'	<i>kamn-i</i>	'stones'
<i>nemec</i>	'German'	<i>nemc-y</i>	'Germans'
<i>nogot'</i>	'nail'	<i>nogt-i</i>	'nails'

Allomorphy

Phonological allomorphy

- Allomorphs that are phonologically similar are **phonological allomorphs**: small differences in the shapes of morphemes that can be regarded as mere differences in pronunciation
- Formal relation between two or more phonological allomorphs: **alternation**
- Each phonological allomorph is called an **alternant**

Allomorphy

Morphophonological rules

- Alternations described using **morphophonological rules**:
A **morphophonological rule** can manipulate an **underlying representation** under certain conditions and yields a **surface representation** (i.e. what is pronounced)
- E.g. Russian: when the stem is followed by a vowel-initial suffix, the vowel *o/e* is often dropped if it is the last vowel in the stem

Morphophonological rule:

“*o/e* in the final stem syllable disappears when the stem is followed by a vowel-initial suffix”

underlying: [zamok] ‘castle-SG’

application: no

surface: [zamok] ‘castle-SG’

underlying: [zamok-i] ‘castle-PL’

application: yes ([zamok-i] → [zamk-i])

surface: [zamk-i] ‘castle-PL’

Allomorphy

Suppletion

- **Suppletion:** the use of one word as the inflected form of another word when the two words are not cognate (i.e. share different roots)
- Morphemes may also have allomorphs that are not similar (i.e. irregular) in pronunciation: **suppletive allomorphs**
- **Strong suppletion:** allomorphs exhibit no similarity at all

<i>go</i>		<i>wen-t</i>		English
<i>good</i>		<i>bett-er</i>		
<i>čelovek</i>	'human being'	<i>ljud-i</i>	'people'	Russian

- **Weak suppletion:** allomorphs exhibit some similarity, but this cannot be described by phonological rules

<i>buy</i>	[bai]	<i>bough-t</i>	[bɔ:t]
<i>catch</i>	[kætʃ]	<i>caugh-t</i>	[kɔ:t]
<i>teach</i>	[ti:tʃ]	<i>taugh-t</i>	[tɔ:t]

- Note that it is often hard to distinguish between weak suppletive allomorphy and phonological allomorphy

Allomorphy

Types of allomorphy

Phonological allomorphy	Alternation could be described by a rule of pronunciation	English plural [s], [z], [əz]
Weak suppletive allomorphy	Allomorphs exhibit some similarity, but this cannot be described by phonological rules	English <i>buy/bough-t</i> , <i>catch/caugh-t</i> , etc.
Strong suppletive allomorphy	Allomorphs exhibit no similarity at all	English <i>good/bett-er</i> , <i>go/wen-t</i> , etc.

Allomorphy

Phonological conditioning

- Phonological allomorphs typically have **phonological conditioning**: the phonological context determines the choice of allomorph
- Example: English plural *-s*
 - [əz] after a sibilant (i.e. [s], [z], [ʃ], [ʒ], [tʃ], or [dʒ]),
e.g. *face-s*, *maze-s*, *bush-es*, *garage-s*, *church-es*, *badge-s*
 - [s] after a voiceless non-sibilant obstruent, e.g. *cat-s*, *book-s*, *cliff-s*
 - [z] elsewhere, e.g. *bag-s*, *bell-s*, *key-s*

Allomorphy

Morphological conditioning

- Stem suppletion usually has **morphological conditioning**: the morphological context (usually, grammatical function) determines the choice of allomorph
- Example: Spanish verb *ir* 'go'
 - *ir* in the infinitive and future tense
 - *va-* in the present and imperfective past tense
 - *fu-* in the perfective past tense

Allomorphy

Lexical conditioning

- Sometimes the choice of a suppletive affix allomorph is dependent on other properties of the base (e.g. semantic properties): **lexical conditioning**
- Persian plural marking: human nouns *-an*, non-human nouns *-ha* (i.e. a semantic property)

<i>-an</i>	<i>mærd</i>	'man'	<i>mærd-an</i>	'men'
	<i>geda</i>	'beggar'	<i>geday-an</i>	'beggars'
<i>-ha</i>	<i>gorbe</i>	'cat'	<i>gorbe-ha</i>	'cats'
	<i>ettefaq</i>	'incident'	<i>ettefaq-ha</i>	'incidents'

- English past participle suffix *-en*
(speakers must simply learn which verbs take *-en* and not the more common suffix *-ed*)
eat/eat-en, write/writt-en, ride/ridd-en

Allomorphy

Types of conditioning

Phonological conditioning	Choice of allomorphs depends on phonological context	English plural depends on final sound in stem (e.g. cat-s [s] / face-s [əz])
Morphological conditioning	Choice of allomorphs depends on the morphological context	Spanish <i>ir</i> , <i>va-</i> , or <i>fu-</i> , depending on tense
Lexical conditioning	Choice of allomorphs depends on the individual lexical item	English past participle <i>-en/-ed</i> is unpredictable and depends on individual verbs

Summary

- **Morphemes**: smallest meaningful constituents
- **Lexeme**: dictionary words (is, was, were → BE)
- **Word-form**: lexeme + grammatical meanings
 - e.g. LIVE + “third person, singular, present tense” = lives
- **Paradigm**: set of word-forms belonging to a lexeme
- Morphological relationships:
 - **Inflection**: lexeme is inflected for grammatical features, e.g. INSULA + num/gen
 - **Derivation**: e.g. lexeme READER is derived from the lexeme READ
 - **Compounding**: e.g. FIRE + WOOD = FIREWOOD
- **Affixes** attach to **base**: e.g. read-able
- **Allomorph**: Morpheme with different phonological shapes under different circumstances. E.g. dog-s / cat-s

Thank you for your attention